



TITLE:

Cats match voice and face: cross-modal representation of humans in cats (*Felis catus*)

AUTHOR(S):

Takagi, Saho; Arahori, Minori; Chijiwa, Hitomi;
Saito, Atsuko; Kuroshima, Hika; Fujita, Kazuo

CITATION:

Takagi, Saho ...[et al]. Cats match voice and face: cross-modal representation of humans in cats (*Felis catus*). *Animal cognition* 2019, 22(5): 901-906

ISSUE DATE:

2019-09

URL:

<http://hdl.handle.net/2433/243873>

RIGHT:

This is a post-peer-review, pre-copyedit version of an article published in 'Animal cognition'. The final authenticated version is available online at: <https://doi.org/10.1007/s10071-019-01265-2>; The full-text file will be made open to the public on 10 May 2020 in accordance with publisher's 'Terms and Conditions for Self-Archiving'. This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ください。

Cats match voice and face: cross-modal representation of humans in cats (*Felis catus*)

Saho Takagi^{1,3}

Minori Arahori^{1,3}

Hitomi Chijiwa^{1,3}

Atsuko Saito²

Hika Kuroshima¹

Kazuo Fujita¹

¹**Affiliation:** *Department of Psychology, Graduate School of Letters, Kyoto*

University

¹**Affiliation address:** *Yoshida-honmachi, Sakyo, Kyoto, 606-8501, Japan.*

²**Affiliation:** *Department of Psychology, Faculty of Human Sciences, Sophia*

University

²**Affiliation address:** *7-1, Kioicho, Chiyoda-ku, Tokyo, 102-8554, Japan*

³**Affiliation:** *Japan Society for the Promotion of Science*

³**Affiliation address:** *5-3-1, Chiyoda-ku, Tokyo, 102-0083, Japan.*

Corresponding author: Saho Takagi, takagi.saho.67x@st.kyoto-u.ac.jp

Telephone: +81 75-753-2442

21 Abstract

22 We examined whether cats have a cross-modal representation of humans, using a
23 cross-modal expectancy violation paradigm originally used with dogs by Adachi et
24 al. (2007). We compared cats living in houses and in cat cafés to assess the
25 potential effect of postnatal experience. Cats were presented with the face of either
26 their owner or a stranger on a laptop monitor after playing back the voice of one of
27 two people calling the subject's name. In half of the trials the voice and face were of
28 the same person (congruent condition) whereas in the other half of trials the
29 stimuli did not match (incongruent condition). The café cats paid attention to the
30 monitor longer in incongruent than congruent conditions, showing an expectancy
31 violation. By contrast, house cats showed no similar tendency. These results show
32 that at least café cats can predict their owner's face upon hearing the owner's voice,
33 suggesting possession of cross-modal representation of at least one human. There
34 may be a minimal kind or amount of postnatal experiences that lead to formation
35 of a cross-modal representation of a specific person.

36 Keywords: Cross-modal representation, Cats, *Felis catus*, Expectancy violation
37 method

38 Introduction

39 Integration of multi - sensory information facilitates the detection or
40 identification of external stimuli. For example, often we hear someone's voice
41 calling us, but we cannot see the person. In this situation we can recall the person's
42 face. This shows that we have a mental representation that integrates information
43 from visual and auditory modalities (cross-modal representation) (see Campanella
44 and Belin 2007 for review). In humans this ability emerges early in life. Bahrick et
45 al. (2005) reported that even 4- to 6-month-old infants perceived face - voice
46 relations of unfamiliar adults.

47 Nonhuman animals also have cross-modal representation of others. This
48 should be an important ability especially for social animals living in complex
49 societies; allowing them to identify individuals, avoid conflicts and maintain social
50 balance, rank, and perhaps cooperation. Some social species are known to have
51 cross-modal representations of conspecifics (chimpanzees (*Pan troglodytes*): Kojima
52 et al. 2003, rhesus macaques (*Macaca mulatta*): Adachi and Hampton 2011; Sliwa
53 et al. 2011, Grey-Cheeked mangabeys (*Lophocebus albigena*): Bovet and Deputte
54 2009, horses (*Equus caballus*): Proops et al. 2009, lions (*Panthera leo*): Gilfillan et
55 al. 2016, goats (*Capra hircus*): Pitcher et al. 2017, crows (*Corvus macrorhynchos*):
56 Kondo et al. 2012). Furthermore, rhesus monkeys, squirrel monkeys (*Saimiri*

57 *boliviensis*) and dogs (*Canis familiaris*) can also form cross-modal representation of
58 familiar members of at least one other species, namely humans (Adachi et al. 2007;
59 Adachi and Fujita 2007; Sliwa et al. 2011).

60 Adachi and Fujita (2007) reported that squirrel monkeys responded differently
61 depending on the familiarity of the human stimuli, using a symbolic matching-to-
62 sample task. They trained monkeys to match photographs of two caretakers and a
63 symbolic visual stimulus. One caretaker was a primary caretaker, more familiar
64 than the other (secondary caretaker). In test trials, a voice which belonged to either
65 the primary or secondary caretaker was played back immediately after the visual
66 sample stimulus disappeared, then two comparison stimuli appeared, one of which
67 the monkey was required to choose. The authors predicted that congruency
68 between voice and sample stimulus would affect matching accuracies. Results
69 showed that accuracies did not differ between congruent and incongruent trials
70 when the primary caretaker's face was the sample, but accuracies were higher in
71 congruent than incongruent trials when the secondary caretaker's face was the
72 sample. Thus, the secondary caretaker's voice did not interfere with matching the
73 primary caretaker's face to the symbolic stimulus, whereas the primary caretaker's
74 voice interfered with matching the secondary caretaker's face to the corresponding

symbolic stimulus. Thus, familiarity of the specific person affected the monkeys' cross-modal representation.

Adachi et al. (2007) reported that pet dogs have a cross-modal representation of their owner. Dogs were presented with a photo of either their owner's or a stranger's face on a monitor after a voice calling subject's name was played back. The voice and face matched in half of the trials and mismatched in the other half. Results showed that dogs looked at the photo longer in both incongruent conditions, suggesting that they predicted the owner's face upon hearing the owner's voice, and another face upon hearing a stranger's voice. Conceivably, extensive experience with a specific person strengthens the formation of such cross-modal representations. Do other companion animals show the same tendency as dogs?

Like dogs, cats are a popular companion animal for humans, and recent studies have shown that like dogs, cats also have remarkable social cognitive abilities. They respond to human pointing cues (Miklósi et al. 2005) and gaze cues (Pongrácz et al. 2018), discriminate human emotional expressions (Galvan and Vonk 2016) and human attentional states (Ito et al. 2016), and refer to human facial expressions in the presence of a mildly frightening object (Merola et al.

93 2015).Saito and Shinozuka (2013), using a habituation-dishabituation procedure,
94 reported that cats discriminated their owner's voice from a stranger's voice.
95 However, it is unknown whether they predict their owner's face after hearing the
96 owner's voice, as expected if integration of the relevant audio-visual information
97 occurs.

98 Here we asked whether cats (*Felis catus*) have cross-modal representations of
99 their owners, using the task originally used with dogs in Adachi et al. (2007). If
100 familiarity of the person affects cross-modal representation, as seen in previous
101 studies, the rearing environment should affect formation of a cross-modal
102 representation of the owner. More specifically, house cats – with a closer
103 relationship with their owner – should show stronger results than cats living at a
104 cat café where many people interact with them each day. In previous research data
105 from these two groups of cats analyzed separately their responses to human voices
106 were different (Saito et al. 2019). The expectancy violation-based prediction was
107 that if cats have a cross-modal representation of their owner they should pay
108 attention to the monitor for longer in incongruent (mis-matching) conditions than
109 congruent (matching) conditions.

110

Methods

Subjects

Eighty-seven domestic cats (*Felis catus*) (48 males, 39 females) participated. Forty-three were kept at five “cat cafés” (24 males, 19 females, mean age 4.14 years, $SD = 2.98$ years, range 4 months to 10.7 years), where many unfamiliar visitors have contact with the cats. There are various types of cat cafes in Japan. Some serve both as a normal cat café where visitors can enjoy interacting with cats and consider fostering a cat. Cats leave these cafés when they find a foster family. We tested cats in cafes where the cats were permanent residents and where they spend all their time. The remaining subjects were house cats (24 males, 20 females, mean age 5.14 years, $SD = 3.18$ years, range 8 months to 12.4 years). Subjects had been with their owner for at least for 4 months in cat cafés and 11 months in households. An additional 23 cats (12 cats from cat cafés and 11 from households) were excluded due to camera error (3 cats), fear (3), or failure to look at the stimuli (no look in all 4 test trials) (17). In addition to approval from the institutional animal experiment committee (see paragraph on compliance with ethical standards), informed consent was obtained from all owners. Cats were not deprived of water or food during the study.

130 Apparatus & Stimuli

131 The auditory stimuli consisted of a recording of either the owner or a same-sex
132 unfamiliar person (stranger) calling the subject's name once. Each owner was
133 instructed to call out the cat's name as they normally would; the stranger was
134 instructed to call out the name in the way the owner did. We recorded the calls
135 using a handheld digital audio recorder (Roland EDIROL R-09, Japan) in WAV
136 format. The sampling rate was 44,100 Hz and the sampling resolution was 16-bit.
137 We used 1-s call stimuli regardless of the cats' names; all voices were adjusted to
138 the same volume with the help of version 2.3.0 of Audacity(R) recording and editing
139 software (Audacity Team 2018). Voice stimuli were played from a speaker (Sanwa
140 MM-SPS2UBK, Japan) connected to a laptop personal computer (NEC Lavie G
141 type Z, Japan) which controlled all experimental stimuli. The visual stimuli
142 consisted of a photo of the face of either the owner or a stranger. We took a digital,
143 full-face, color photo of each person smiling, and stored the photo in PNG format.
144 Presented photos were ca. 16.5 x 16 cm on the 13.3 -in. monitor of the laptop
145 computer. The background was always black.

146 The test was recorded by three video cameras (JVC GZ-E565-R, Japan; SONY
147 HDR-CX390, Japan; SONY HDR-CX675, Japan), one placed in front of subject,

148 another placed slightly to one side, and the other placed behind subject; all
149 cameras focused on the cat.
150
151 Procedure
152 Cats were tested individually in their familiar place: house or café. Before testing
153 we waited until cats appeared relaxed in the presence of the experimenter; this
154 took about 15 min for house cats whereas almost all café cats ended no such
155 familiarization time. An experimenter gently restrained the cat on the floor in front
156 of the laptop computer, about 40 cm away. The experimenter started a trial by
157 pressing a key on the computer when the subject was looking toward the monitor.
158 Each trial consisted of two phases: the voice phase and the face phase. In the voice
159 phase, one stimulus voice was played back from the speakers linked to the laptop
160 every 1 s, for a total of four presentations. Immediately after the fourth auditory
161 stimulus either the owner's or a stranger's face appeared on the monitor for 7-s
162 (face phase) (see Fig.1). The experimenter restrained the cat throughout the voice
163 phase and released it at the start of the face phase; some cats stayed, whereas
164 others moved around to explore the monitor. A trial ended when the face on the
165 monitor disappeared.

166 There were four experimental conditions according to the combination of face
167 and voice: owner-congruent, owner-incongruent, stranger-congruent, and stranger-
168 incongruent. For example, in the owner-incongruent condition, a stranger's voice
169 was played in the voice phase, but the owner's face appeared in the face phase.
170 These four trials were presented in pseudo-random order with the restriction that
171 the same voice was never repeated on consecutive trials.

172 Our hypothesis was that cats would pay attention to the face (the monitor)
173 longer in the incongruent condition than the congruent condition. Each subject
174 received four trials in a single session, with an inter-trial interval of at least 3 min.
175 For ethical reasons we immediately halted the procedure if the subject refused to
176 be placed in front of the monitor; three subjects participated in only the first trial,
177 three in the first and second trials, while four subjects received no fourth trial.
178 During the interval, cats acted freely in the experimental room. The experimenter
179 restraining the cat was ignorant of the condition; she closed her eyes during the
180 test trials and avoided making eye contact with the subject. Presentation of voice
181 and face stimuli was controlled via a Visual Studio 2013 program on the laptop
182 personal computer.

183

184 Analysis

185 A coder (H.C.), blind to the conditions, counted the number of frames (30
186 frames/1 sec.) in which cats paid attention to the monitor in the face phase (total 7
187 sec) for each condition. Paying attention was defined as looking at or sniffing the
188 monitor. We could not discriminate these two acts because a few cats did not touch
189 the monitor with their nose while sniffing. Trials in which subject did not look at
190 the monitor at all were excluded from the analyses because we could not know if
191 expectancy violation occurred. Sixty-four trials were excluded for café cats, 65 for
192 house cats (no significant difference; Fisher's Exact Test: $p = .73$). Table 1 shows
193 valid data points, i.e., the number of trials cats looked at the monitor in each
194 condition. The videos were analyzed using Adobe Premiere CS6 (USA) software.

195 To check the reliability of coding, an assistant who was blind to the conditions
196 coded a randomly chosen 20% of the videos. The correlation between the two coders
197 was excellent for time spent paying attention to the monitor (Pearson's $r = 0.97$, n
198 $= 40$, $p < 0.01$).

199 All statistical analyses were conducted with R version 3.5.1 (R Core Team,
200 2018). Attention to the monitor was analyzed by a linear mixed model (LMM) using
201 a lmer function in lme4 package version 1.1.10 (Bates Martin Bolker and Walker

2015), in which face (owner/stranger), congruency (congruent/incongruent), home environment (café/house) and an interaction between congruency and home environment were entered as fixed factors and subject identity was entered as a random factor. To test whether effects of factor were significant, we ran F tests by an Anova function in car package (Fox et al 2012). We used a diffIsmeans function in lmerTest package (Kuznetsova Brockhoff and Christensen 2017) which tested differences of least squares means to compare each condition. Degrees of freedom were adjusted by Kenward-Roger and p-value was adjusted by the Holm procedure.

Results

Fig. 2 shows time spent paying attention to the monitor during the face phase in café cats (A) and house cats (B). Contrary to our prediction, café cats showed more attention to the monitor in both incongruent conditions, whereas house cats showed no clear tendency; they attended to the monitor almost randomly. LMM revealed that significant main effects of congruency ($F(1, 60.87) = 4.10, p = .04$), home environment ($F(1, 79.03) = 8.06, p < .01$), and an interaction between congruency and home environment ($F(1, 60.71) = 7.76, p < .01$). There was no significant main effect of face ($F(1, 96.78) = 0.06, p = .79$).

The test of differences of least squares means showed a significant difference

221 between congruent and incongruent conditions in café cats ($p < .01$), between café
222 cats and house cats in congruent conditions ($p < .01$), and between café cats in
223 congruent conditions and house cats in incongruent conditions ($p < .01$).

224

225 Discussion

226 We used an expectancy violation procedure to ask whether cats have a cross-
227 modal representation of their owner. We presented the face of either the owner or
228 stranger after playing back the voice of the owner or a stranger calling the subjects'
229 name. Results showed that café cats paid attention to the monitor for longer in
230 both incongruent conditions, when voices and face were mismatched, whereas
231 house cats showed no clear trends. These results contradict our prediction and
232 suggest clearly that café cats predict the owner's face upon hearing the
233 corresponding voice, demonstrating a cross-modal representation of a specific
234 person; whether house cats have this cross-modal ability remains to be further
235 examined. The results also indicate that cross-modal representations of others are
236 not exclusive to species that form complex social groups, such as dogs, but also
237 more solitary species, such as cats (Bradshaw 2016).

238 There are three possible explanations for our failure to demonstrate a cross-
239 modal representation in house cats. First, cross-modal representation of a specific

240 person might be affected by factors other than familiarity with that individual.

241 Café cats typically see and interact with multiple strangers on a daily basis. People

242 with greater experience of heterospecific faces can discriminate them better than

243 people with fewer such experiences (Dufour and Petit 2010). Also, older captive

244 chimpanzees discriminated human faces better than younger chimpanzees

245 probably because older chimpanzees had more experiences to see a variety of

246 human faces (Dahl et al 2013). These results raise the possibility that café cats

247 greater experience of a variety of human faces and voices might result in better

248 discrimination abilities.

249 Second, greater experience of seeing and interacting with people might promote

250 application of an “exclusive rule.” Our task required cats to predict a stranger from

251 a stranger’s voice in one of the incongruent conditions. It should be more difficult to

252 predict the stranger from the stranger’s voice than the owner from the owner’s

253 voice using an exclusion rule. Kondo et al (2012) demonstrated that crows did not

254 react even when a familiar crow’s calls were played back followed by an unfamiliar

255 bird presented visually, suggesting that they did not exclusively predict “a

256 stranger”. Similar asymmetrical results were obtained in horses (Proops and

257 McComb 2012). In contrast, café cats showed expectancy violation in both

258 incongruent conditions; they showed no asymmetry. This suggests that the café
259 cats exclusively predicted a non-owner face, using the exclusion rule. Conceivably,
260 increased opportunities to see various people might improve cross-modal
261 representations of others. Home environments that differ from normal pet
262 environments in terms of seeing human strangers might explain why café cats
263 showed the clearer expectancy violation.

264 Finally, house cats might have been more nervous during the test. We asked the
265 owner to remain in another room because we wanted to test cats' representation of
266 their owner. Some cats may not have felt sufficiently at ease in the presence of only
267 the experimenter. More house cats remained immobile after the experimenter
268 released them in the face phase; a freezing reaction might have resulted in longer
269 looking times in all conditions compared to café cats. To exclude such a possibility
270 future work should use a more natural experimental setting less likely to cause
271 stress in house cats, or conduct a test that objectively estimates their stress level.

272 One may argue that house cats did not discriminate between the owner's face
273 and a stranger's face, given their lack of differential responses across conditions.
274 However, previous studies have shown that house cats respond differently to
275 familiar and unfamiliar humans facing them directly (Collard 1967; Ellis

276 Thompson Guijarro and Zulch 2015; Galvan and Vonk 2016). Further research
277 should be conducted on cats' ability to discriminate the owner's face from a
278 stranger's face when only visual information is presented.

279 We used voice and face to examine cats' cross-modal recognition of humans.
280 However, cats also use their olfactory sense to recognize others (Gorman and
281 Trowbridge 1989). Further research should examine whether olfactory information
282 is also integrated in cross-modal representations of others.

283 Cross-modal recognition is not limited to a one-to-one relation (owner's voice -
284 face) as in this study. For example, dogs can show more general cross-modal
285 recognition: Taylor Reby and McComb (2011) examined whether dogs could match
286 frequency of growls and dogs' body size. Dogs spent more time looking at a correct
287 model (small body - high frequency, or big body - low frequency) than an incorrect
288 model (small body - low frequency, or big body - high frequency), suggesting that
289 they relate information about body size to "voices." Furthermore, dog cross-modally
290 matched a human male or female voice and a male or female face (Takaoka et al
291 2013). It is still unknown whether cats have similar cross-modal recognition
292 abilities beyond one-to-one correspondence; this is another issue for future study.
293

294 Acknowledgments

295 This study was financially supported by the Grant-in-aid for Scientific
296 Research (KAKENHI) No. 17J08974 to S. Takagi, No. JP16J1034 to M. Arahori,
297 No. JP16J08691 to H. Chijiiwa, No. 25118003 to A. Saito and Nos. 25240020,
298 26119514, 16H01505, 15K12047, 25118002, and 16H06301 to K. Fujita from the
299 Japan Society for the Promotion of Science (JSPS). The authors acknowledge with
300 thanks all owners and cats who volunteered in this study. The authors also wish to
301 thank Dr. James R. Anderson for editing the article.

302

303 Compliance with ethical standards

304 This study adhered to the ethical guidelines of Kyoto University, and was
305 approved by the Animal Experiment Committee of the Graduate School of Letters,
306 Kyoto University.

307

308 Competing interests

309 The authors declare no conflicts of interest.

310

311 Reference

- 312 Adachi, I., & Fujita, K. (2007). Cross-modal representation of human caretakers in
313 squirrel monkeys. *Behavioural Processes*, 74, 27-32.
- 314 Adachi, I., & Hampton, R. R. (2011). Rhesus monkeys see who they hear:
315 spontaneous cross-modal memory for familiar conspecifics. *PLoS One*, 6,
316 e23345.
- 317 Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon
318 hearing the owner's voice. *Animal Cognition*, 10, 17-21.
- 319 Audacity Team (2018). Audacity(R): Free Audio Editor and Recorder [Computer
320 application]. Version 2.3.0 retrieved December 20th 2018 from
321 <https://audacityteam.org/> .
- 322 Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant
323 learning about specific face-voice relations. *Developmental Psychology*, 41,
324 541-552.
- 325 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-
326 Effects Models Using lme4. *Journal of Statistical Software*, 67, 1-48.
- 327 Bovet, D., & Deputte, B. L. (2009). Matching vocalizations to faces of familiar
328 conspecifics in grey-cheeked mangabeys (*Lophocebus albigena*). *Folia*
329 *Primatologica*, 80, 220-232.

- 330 Bradshaw, J. W. (2016). Sociality in cats: A comparative review. *Journal of*
331 *Veterinary Behavior: Clinical Applications and Research*, 11, 113-124.
- 332 Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception.
333 *Trends in Cognitive Sciences*, 11, 535-543.
- 334 Collard, R. R. (1967). Fear of strangers and play behavior in kittens with varied
335 social experience. *Child Development*, 877-891.
- 336 Dahl, C. D., Rasch, M. J., Tomonaga, M., & Adachi, I. (2013). Developmental
337 processes in face perception. *Scientific Reports*, 3, 1044.
- 338 Ellis, S. L. H., Thompson, H., Guijarro, C., & Zulch, H. E. (2015). The influence of
339 body region, handler familiarity and order of region handled on the domestic
340 cat's response to being stroked. *Applied Animal Behaviour Science*, 173, 60-67.
- 341 Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., & Heiberger,
342 R. (2012). Package 'car'. Vienna: R Foundation for Statistical Computing.
- 343 Galvan, M., & Vonk, J. (2016). Man's other best friend: domestic cats (*F. silvestris*
344 *catus*) and their discrimination of human emotion cues. *Animal Cognition*, 19,
345 193-205.
- 346 Gilfillan G, Vitale J, McNutt JW, McComb K. (2016). Cross-modal individual
347 recognition in wild African lions. *Biology Letters*, 12, 20160323.

- 348 Gorman, M. L., & Trowbridge, B. J. (1989). The role of odor in the social lives of
349 carnivores. In *Carnivore behavior, ecology, and evolution* (pp. 57-88). Springer,
350 Boston, MA.
- 351 Ito, Y., Watanabe, A., Takagi, S., Arahori, M., & Saito, A. (2016). Cats beg for food
352 from the human who looks at and calls to them: Ability to understand humans'
353 attentional states. *Psychologia*, 59, 112-120.
- 354 Kojima, S., Izumi, A., & Ceugniet, M. (2003). Identification of vocalizers by pant
355 hoots, pant grunts and screams in a chimpanzee. *Primates*, 44, 225-230.
- 356 Kondo, N., Izawa, E. I., & Watanabe, S. (2012). Crows cross-modally recognize
357 group members but not non-group members. *Proceedings of the Royal Society
358 of London B: Biological Sciences*, 279, 1937-1942.
- 359 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package:
360 tests in linear mixed effects models. *Journal of Statistical Software*, 82.
- 361 Merola, I., Lazzaroni, M., Marshall-Pescini, S., & Prato-Previde, E. (2015). Social
362 referencing and cat-human communication. *Animal Cognition*, 18, 639-648.
- 363 Miklósi, Á., Pongrácz, P., Lakatos, G., Topál, J., & Csányi, V. (2005). A comparative
364 study of the use of visual communicative signals in interactions between dogs

- 365 (*Canis familiaris*) and humans and cats (*Felis catus*) and humans. *Journal of*
366 *Comparative Psychology*, 119, 179-186.
- 367 Pitcher, B. J., Briefer, E. F., Baciadonna, L., & McElligott, A. G. (2017). Cross-
368 modal recognition of familiar conspecifics in goats. *Royal Society Open Science*,
369 4, 160346.
- 370 Pongrácz, P., Szapu, J. S., & Faragó, T. (2018). Cats (*Felis silvestris catus*) read
371 human gaze for referential information. *Intelligence*, in press.
- 372 Proops, L., & McComb, K. (2012). Cross-modal individual recognition in domestic
373 horses (*Equus caballus*) extends to familiar humans. *Proceedings of the Royal*
374 *Society of London B: Biological Sciences*, rspb20120626.
- 375 Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in
376 domestic horses (*Equus caballus*). *Proceedings of the National Academy of*
377 *Sciences*, 106, 947-951.
- 378 R Core Team (2018). R: A language and environment for statistical computing. R
379 Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-](https://www.R-project.org/)
380 [project.org/](https://www.R-project.org/).
- 381 Saito, A., & Shinozuka, K. (2013). Vocal recognition of owners by domestic cats
382 (*Felis catus*). *Animal Cognition*, 16, 685-690.

- 383 Saito, A., Shinozuka, K., Ito, Y., & Hasegawa, T. (2019) Domestic cats (*Felis catus*)
384 discriminate their names from other words, *Scientific Reports*, in press.
- 385 Sliwa, J., Duhamel, J. R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice–face
386 identity matching by rhesus monkeys for familiar conspecifics and humans.
387 *Proceedings of the National Academy of Sciences*, 108, 1735-1740.
- 388 Takagi, S., Arahori, M., Chijiwa, H., Tsuzuki, M., Hataji, Y., & Fujita, K. (2016).
389 There's no ball without noise: cats' prediction of an object from noise. *Animal*
390 *Cognition*, 19, 1043-1047.
- 391 Takaoka, A., Morisaki, A., & Fujita, K. (2013). [in Japanese with English abstract]
392 Cross-modal concept of human gender in dogs (*Canis familiaris*). *The Japanese*
393 *Journal of Animal Psychology*, 63, 123-130.
- 394 Taylor, A. M., Reby, D., & McComb, K. (2011). Cross modal perception of body size
395 in domestic dogs (*Canis familiaris*). *PLoS One*, 6, e17069.
- 396

397

398 **Legend**

399 **Fig. 1**

400 **Fig.1 Diagram illustrating each condition. Face was presented in the monitor (Face**
401 **phase) immediately after voices was played back (Voice phase). The face and voice**
402 **matched in half of the trials (congruent condition) whereas they mismatched in the**
403 **other half of trials (incongruent condition). Black line represents Congruent**
404 **conditions, dotted line represents Incongruent conditions.**

405

406 **Fig.2**

407 **Time spent paying attention to the monitor in (A) Café cats and (B) House cats in**
408 **the Face phase. White bar represents congruent conditions, Black bar represents**
409 **incongruent conditions. Error bar indicates SE. Unit of Y axis is frames (30 frames/1**
410 **sec.).**

411

412

413 **Table. 1 The number of valid data points representing the number of trials cats**
414 **looked at the monitor in each condition.**

Figure 1

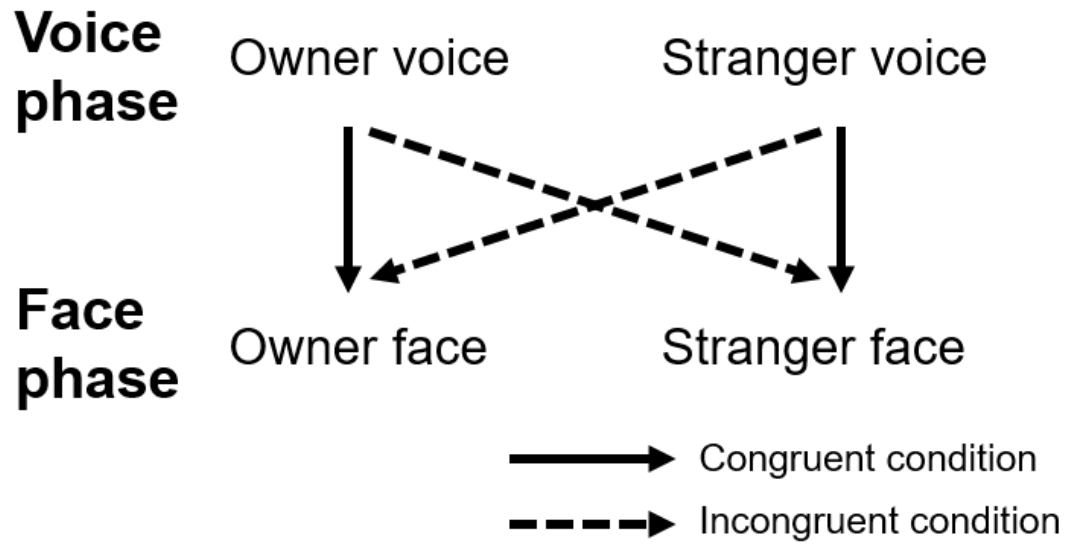


Figure 2

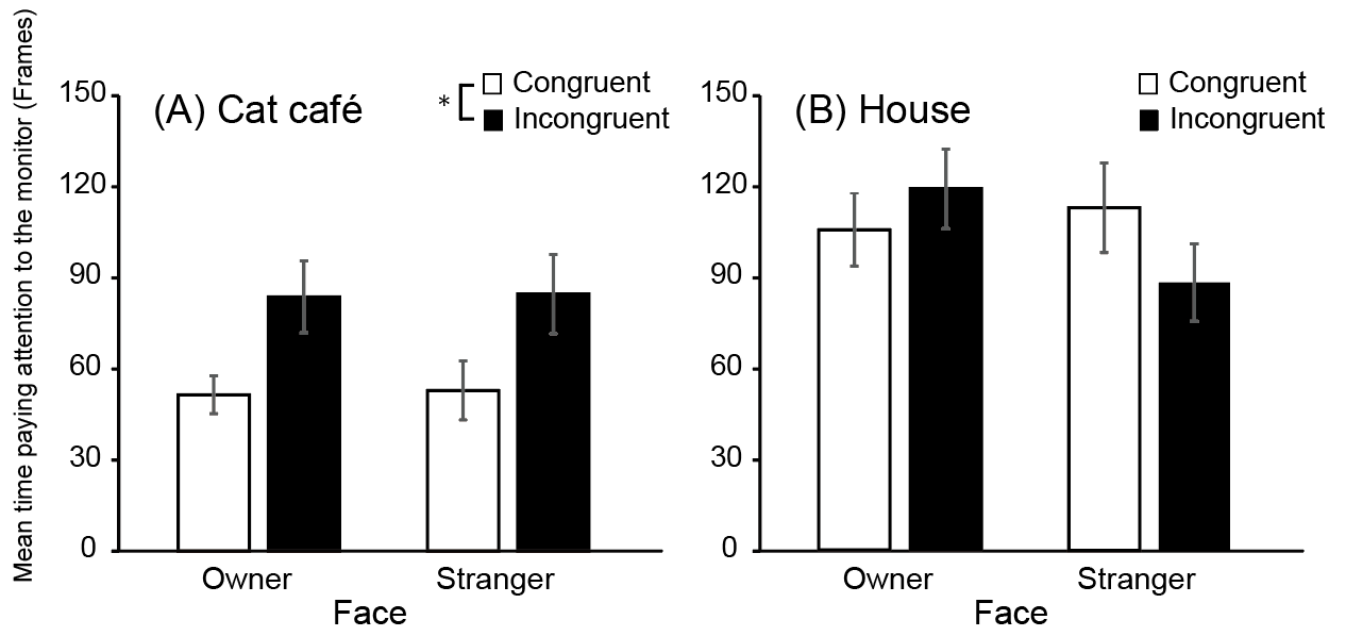


Table.1

Face	Owner		Stranger	
	Congruent	Incongruent	Congruent	Incongruent
Café	22	28	23	22
House	29	23	28	25